

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平8-328583

(43)公開日 平成8年(1996)12月13日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 L 3/00 5/06	5 6 1		G 1 0 L 3/00 5/06	5 6 1 A D

審査請求 有 請求項の数 3 O L (全 4 頁)

(21)出願番号 特願平7-136725

(22)出願日 平成7年(1995)6月2日

(71)出願人 000004237

日本電気株式会社
東京都港区芝五丁目7番1号

(72)発明者 坂井 信輔

東京都港区芝五丁目7番1号 日本電気株
式会社内

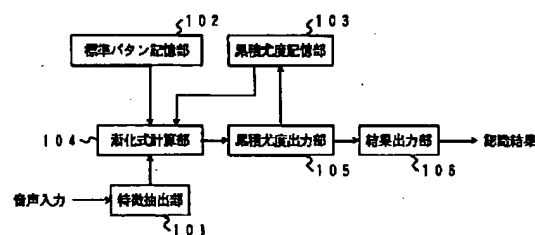
(74)代理人 弁理士 後藤 洋介 (外2名)

(54)【発明の名称】 音声認識装置

(57)【要約】

【目的】 周囲の雑音環境の変動やビーム幅の設定値の変化に対して安定した探索効率を有しながら、刈り取り閾値決定のための処理量が小さい、高速な音声認識装置を提供する。

【構成】 特徴抽出部101は、音声入力の特徴ベクトルの時系列に変換し、漸化式計算部104に出力する。標準ボタン記憶部102は、標準ボタンを記憶する。累積尤度記憶部103は、累積尤度出力部105から出力される累積尤度を記憶する。漸化式計算部104は、第1フレームの特徴ベクトル、標準ボタン、及び第1-1フレームまでの累積尤度から第1フレームまでの累積尤度を求める。累積尤度出力部105は、入力された累積尤度の集合から、あるKフレームの部分区間では、第M位までの候補を選択、それ以降は、Kフレームの部分区間で求めた最尤候補と第M位の候補との累積尤度差の平均値を用いて候補を選択し、累積尤度記憶部103に出力する。結果出力部106は、最終フレームまでの累積尤度に基づいて認識結果を出力する。



【特許請求の範囲】

【請求項1】 入力音声の最初の数フレームで、最大尤度と一定個数番目の候補の尤度との差をもとめておき、それ以後は、前記差を用いて、ビームサーチにおける候補刈り取りのための閾値を設定することを特徴とする音声認識装置。

【請求項2】 音声信号を分析して特徴ベクトル時系列を出力する特徴抽出部と、

あらかじめ作成された標準ボタンを蓄えておく標準ボタン記憶部と、

累積尤度を保持する累積尤度記憶部と、

前記累積尤度記憶部に蓄えられた累積尤度と前記特徴ベクトルの時系列と前記標準ボタンとから新しい累積尤度を求める漸化式計算部と、

前記特徴ベクトル時系列のうち、ある部分系列に対しては、漸化式計算部でもとめられた累積尤度のうち一定個数を出力するとともに、最大の累積尤度と前記一定個数番目の尤度との差を蓄積しておき、それ以降の部分系列に対しては、前記蓄積された尤度の差を用いて求められた閾値により、出力する累積尤度を決定する累積尤度出力部と、

前記累積尤度出力部から出力される累積尤度より前記音声信号に対する認識結果を求める結果出力部とを有することを特徴とする音声認識装置。

【請求項3】 入力音声の任意の個数の部分系列のおおのに対して、第M位の候補の累積尤度の最大累積尤度との差の平均値を求め、次の部分系列の間では、前の部分系列で求めた前記差の平均値を用いて候補刈り取りの閾値を設定することを特徴とする音声認識装置における閾値設定方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、音声認識装置に関する。

【0002】

【従来の技術】 音声認識装置は非常に大きい演算量を必要とするため、従来よりビームサーチによる演算量の削減が試みられている。ビームサーチによる候補刈り取りのためのビームの幅の設定法としては、各候補刈り取り時に、尤度の高いものから一定数の候補を残す方法と、最大尤度から一定幅の範囲の尤度をもつ候補を残す方法が良く知られている。

【0003】 伊藤らによる、音響学会研究発表会講演論文集1993年10月73～74ページに掲載の論文「連続音声認識におけるビームサーチ」においては、一定数の候補を残す方法のほうが、ビーム幅の設定値の変化に対して探索効率が安定していると報告されている。また、最大尤度から一定幅の範囲の尤度をもつ候補の数は、発声時の周囲雑音環境の影響を受けて変動すると考えられるが、一定数の候補を残す方法においては、その

ような候補数の変動はない。

【0004】

【発明が解決しようとする課題】 しかしながら、上述の一定数（仮にM個とする）の候補を残す方法では、候補刈り取り時に第M位の候補を求めるための並べ替え処理が必要となるために、処理量が多いという欠点があった。

【0005】

【課題を解決するための手段】 請求項1記載の発明によれば、入力音声の最初の数フレームで、最大尤度と一定個数番目の候補の尤度との差をもとめておき、それ以後は、前記差を用いて、ビームサーチにおける候補刈り取りのための閾値を設定することを特徴とする音声認識装置が得られる。

【0006】 請求項2記載の発明によれば、音声信号を分析して特徴ベクトル時系列を出力する特徴抽出部と、あらかじめ作成された標準ボタンを蓄えておく標準ボタン記憶部と、累積尤度を保持する累積尤度記憶部と、前記累積尤度記憶部に蓄えられた累積尤度と前記特徴ベクトルの時系列と前記標準ボタンとから新しい累積尤度を求める漸化式計算部と、前記特徴ベクトル時系列のうち、ある部分系列に対しては、漸化式計算部でもとめられた累積尤度のうち一定個数を出力するとともに、最大の累積尤度と前記一定個数番目の尤度との差を蓄積しておき、それ以降の部分系列に対しては、前記蓄積された尤度の差を用いて求められた閾値により、出力する累積尤度を決定する累積尤度出力部と、前記累積尤度出力部から出力される累積尤度より前記音声信号に対する認識結果を求める結果出力部とを有することを特徴とする音声認識装置が得られる。

【0007】 請求項3記載の発明によれば、入力音声の任意の個数の部分系列のおおのに対して、第M位の候補の累積尤度の最大累積尤度との差の平均値を求め、次の部分系列の間では、前の部分系列で求めた前記差の平均値を用いて候補刈り取りの閾値を設定することを特徴とする音声認識装置における閾値設定方法が得られる。

【0008】

【実施例】 次に、本発明について図面を参照して説明する。

【0009】 図1は、本発明の一実施例を示すブロック図である。図1を参照すると本発明の実施例は、特徴抽出部101と、標準ボタン記憶部102と、累積尤度記憶部103と、漸化式計算部104と、累積尤度出力部105と、結果出力部106とから構成される。

【0010】 特徴抽出部101は、音声入力の特徴ベクトルの時系列に変換し、漸化式計算部104に出力する。標準ボタン記憶部102は、標準ボタンを記憶する。累積尤度記憶部103は、累積尤度出力部105から出力される累積尤度を記憶する。処理が開始される以前には、全認識パス候補に対して累積尤度の初期値1、

3

0を保持する。漸化式計算部104は、第1フレームの特徴ベクトル、標準パターン、および第1-1フレームまでの累積尤度から、第1フレームまでの累積尤度を求める。累積尤度出力部105は、入力された累積尤度の集合から、次フレームの累積尤度計算に用いられるものを選択し、累積尤度記憶部103に出力する。結果出力部106は、最終フレームまでの累積尤度に基づいて認識結果を出力する。

【0011】次に、図1及び図2を参照して、本実施例の動作について説明する。

【0012】入力された音声は、特徴抽出部101において、一定の時間間隔ごとに、音声の周波数をスペクトルをあらわす特徴ベクトルに変換され、漸化式計算部104に出力される。この一定の時間間隔を以下ではフレームと呼ぶ。第1フレームにおいて、漸化式計算部104では、標準パターン記憶部102に保持されている標準パターン

$$g_v(i, j) = \max_{x=0,1} (g_v(i-1, j-x) + l(i, j))$$

累積尤度出力部105は、あらかじめ決められたKと比較して、 $1 \leq K$ であるならば、最大値から第M番目の累積尤度を求め、これを候補刈り取りのための閾値 θ とし、これと最大尤度との差dを求める。後で平均を求めるために、dの累積値 S_d を、 $S_d = S_d + d$ と更新する(ステップ2、6、及び7)。

【0014】なお、 S_d は、第1フレーム以前には0に初期化しておく。

【0015】 $i=K$ の場合は、Kフレーム間の最大尤度と候補刈り取り閾値との差の平均 $D = S_d / K$ を求める(ステップ3)。

【0016】また、 $i > K$ の場合は、候補刈り取りのための閾値 θ は、 $\theta = g_{\dots} - D$ とする。 g_{\dots} は、第1フレームにおける累積尤度の最大値である(ステップ4)。

【0017】各フレームにおいて、累積尤度出力部105は、累積尤度の閾値 θ よりも大きい尤度をもつ認識パス候補のみを累積尤度記憶部に出力する(ステップ5)。

【0018】現フレームが最終フレームである場合は、累積尤度出力部105は、標準パターンの終端点に達したすべての認識パス候補を結果出力部106に出力する。結果出力部106は、累積候補が最大の認識パス候補をもとめ、認識結果を出力する(ステップ8、9)。

【0019】以上、本実施例では、入力音声の最初のKフレームで、第M位の候補の累積尤度の最大累積尤度との差の平均値を求めるという例によって説明したが、さらに一般には、入力音声の任意の L_{\dots} 個の部分系列 $l_{i1}, l_{i2}, \dots, l_{iL_{\dots}}$ ($L_{\dots} \geq 1$) (これらの

4

*REF = $\{R_1, \dots, R_N\}$ 、ここで $R_v = \{r_v(1), \dots, r_v(J_v)\}$

を用いて、現在のフレームの特徴ベクトルの各標準パターンに対する局所的尤度

$l_v(i, j)$ ($w=1, \dots, N, j=1, \dots, J_v$)

を求める。ここで、Nは標準パターン数、 J_v はw番目の標準パターンのフレーム長である。次に、この局所的尤度、及び累積尤度記憶部103に保持されている第1-1フレームの累積尤度集合

$$G = \{g_1(i-1, 1), \dots, g_1(i-1, J_1), \dots, g_N(i-1, 1), \dots, g_N(i-1, J_N)\}$$

から、動的計画法に基づいた最大化処理により、下記数1として現在のフレームの認識パス候補およびその累積尤度を求める(図2のステップ1)。

【0013】

【数1】

部分系列を仮に学習区間と呼ぶ)のおおのに対して、上記の平均値を求め、学習区間 l_{i1} と次の学習区間 l_{i+1} の間では、 l_{i1} で求めた差の平均値を用いて候補刈り取りの閾値を設定するという方法をとることができる。

【0020】

【発明の効果】以上説明したように、本発明による音声認識装置は、周囲の雑音環境の変動やビーム幅Mの設定値の変化に対応して第M位の候補の累積尤度と最大の累積尤度の差が大きく変動するような場合でも、入力の一部を用いてこの差の平均値を求めておき、これを用いて刈り取り閾値の決定を行なうので、入力 of 全ての区間に対して累積尤度第M位までの候補を残す方法に準ずる候補の刈り取りが行なわれ、安定した探索効率を有しながらも、刈り取り閾値決定のための処理量が多くならないという効果を有する。

【図面の簡単な説明】

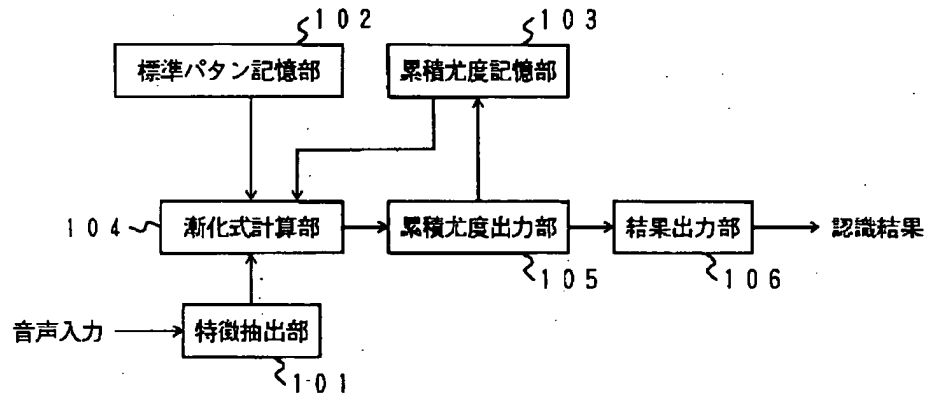
【図1】本発明の音声認識装置の一実施例の構成を示したブロック図である。

【図2】図1に示す音声認識装置の一実施例の処理の流れを示したフローチャートである。

【符号の説明】

- 101 特徴抽出部
- 102 標準パターン記憶部
- 103 累積尤度記憶部
- 104 漸化式計算部
- 105 累積尤度出力部
- 106 結果出力部

【図1】



【図2】

